IN VITRO SELECTION OF PROTEINS USING RNA-PROTEIN FUSIONS

Background of the Invention

This invention relates to protein selection methods.

The invention was made with government support under grant F32 GM17776-01 and F32 GM17776-02. The government has certain rights in the invention.

Methods currently exist for the isolation of RNA and DNA molecules based on their functions. For example, experiments of Ellington and Szostak (Nature **346**:818 (1990); and Nature **355**:850 (1992)) and Tuerk and Gold (Science **249**:505 (1990); and J. Mol. Biol **222**:739 (1991)) have demonstrated that very rare (i.e., less than 1 in 10^{13}) nucleic acid molecules with desired properties may be isolated out of complex pools of molecules by repeated rounds of selection and amplification. These methods offer advantages over traditional genetic selections in that (i) very large candidate pools may be screened (> 10^{15}), (ii) host viability and in vivo conditions are not concerns, and (iii) selections may be carried out even if an in vivo genetic screen does not exist. The power of in vitro selection has been demonstrated in defining novel RNA and DNA sequences with very specific protein binding functions (see, for example, Tuerk and Gold, Science **249**:505 (1990); Irvine et al., J. Mol. Biol **222**:739 (1991); Oliphant et al., Mol. Cell Biol. **9**:2944 (1989); Blackwell et al., Science **250**:1104 (1990); Pollock and Treisman, Nuc. Acids Res. **18**:6197 (1990); Thiesen and Bach, Nuc. Acids Res. **18**:3203 (1990); Bartel et al., Cell **57**:529 (1991); Stormo and Yoshioka, Proc. Natl. Acad. Sci. (USA) **88**:5699 (1991); and Bock et al., Nature **355**:564 (1992)), small molecule binding functions (Ellington and Szostak, Nature **346**:818 (1990); Ellington and Szostak, Nature **355**:850 (1992)), and catalytic functions (Green et al., Nature **347**:406 (1990); Robertson and Joyce, Nature **344**:467 (1990); Beaudry and Joyce, Science **257**:635 (1992); Bartel

and Szostak, Science **261**:1411 (1993); Lorsch and Szostak, Nature **371**:31-36 (1994); Cuenoud and Szostak, Nature **375**:611-614 (1995); Chapman and Szostak, Chemistry and Biology **2**:325-333 (1995); and Lohse and Szostak, Nature **381**:442-444 (1996)). A similar scheme for the selection and amplification of proteins has not been demonstrated.

Summary of the Invention

The purpose of the present invention is to allow the principles of in vitro selection and in vitro evolution to be applied to proteins. The invention facilitates the isolation of proteins with desired properties from large pools of partially or completely random amino acid sequences. In addition, the invention solves the problem of recovering and amplifying the protein sequence information by covalently attaching the mRNA coding sequence to the protein molecule.

In general, the inventive method consists of an in vitro transcription/ translation protocol that generates protein covalently linked to the 3' end of its own mRNA, i.e., an RNA-protein fusion. This is accomplished by synthesis and in vitro translation of an mRNA molecule with a peptide acceptor attached to its 3' end. One preferred peptide acceptor is puromycin, a nucleoside analog that adds to the C-terminus of a growing peptide chain and terminates translation. In one preferred design, a DNA sequence is included between the end of the message and the peptide acceptor which is designed to cause the ribosome to pause at the end of the open reading frame, providing additional time for the peptide acceptor (for example, puromycin) to accept the nascent peptide chain.

If desired, the resulting RNA-protein fusion allows repeated rounds of selection and amplification because the protein sequence information may be recovered by reverse transcription and amplification, and may then be transcribed, modified, and in vitro translated to generate mRNA-protein fusions for the next round of selection. The ability to carry out multiple rounds of selection and amplification enables the isolation of very rare molecules, e.g., one desired molecule out of a pool of 10^{15} members. This in

turn allows the isolation of new or improved proteins which specifically recognize virtually any target or which catalyze desired chemical reactions.

Accordingly, in a first aspect, the invention features a method for in vitro selection of a desired protein, involving the steps of: (a) providing a population of candidate RNA molecules, each of which includes a translation initiation sequence and a start codon operably linked to a candidate protein coding sequence and each of which is covalently bonded to a peptide acceptor at the 3' end of the candidate protein coding sequence; (b) in vitro translating the candidate protein coding sequences to produce a population of candidate RNA-protein fusions; and (c) identifying a desired RNA-protein fusion, thereby selecting the desired protein.

In a related aspect, the invention features a method for in vitro selection of a DNA molecule which encodes a desired protein, involving the steps of: (a) providing a population of candidate RNA molecules, each of which includes a translation initiation sequence and a start codon operably linked to a candidate protein coding sequence and each of which is covalently bonded to a peptide acceptor at the 3' end of the candidate protein coding sequence; (b) in vitro translating the candidate protein coding sequences to produce a population of candidate RNA-protein fusions; (c) identifying a desired RNA-protein fusion; and (d) generating from the RNA portion of the fusion a DNA molecule which encodes the desired protein.

In another related aspect, the invention features a method for in vitro selection of a protein having an altered function relative to a reference protein, involving the steps of: (a) producing a population of candidate RNA molecules from a population of DNA templates, the candidate DNA templates each having a candidate protein coding sequence which differs from the reference protein coding sequence, the RNA molecules each comprising a translation initiation sequence and a start codon operably linked to the candidate protein coding sequence and each being covalently bonded to a peptide acceptor at the 3' end; (b) in vitro translating the candidate protein coding sequences to

produce a population of candidate RNA-protein fusions; and (c) selecting an RNA-protein fusion having an altered function, thereby selecting the protein having the altered function.

In yet another related aspect, the invention features a method for in vitro selection of a DNA molecule which encodes a protein having an altered function relative to a reference protein, involving the steps of: (a) producing a population of candidate RNA molecules from a population of candidate DNA templates, the candidate DNA templates each having a candidate protein coding sequence which differs from the reference protein coding sequence, the RNA molecules each comprising a translation initiation sequence and a start codon operably linked to the candidate protein coding sequence and each being covalently bonded to a peptide acceptor at the 3' end; (b) in vitro translating the candidate protein coding sequences to produce a population of RNA-protein fusions; (c) selecting an RNA-protein fusion having an altered function; and (d) generating from the RNA portion of the fusion a DNA molecule which encodes the protein having the altered function.

In preferred embodiments of the above methods, the peptide acceptor is puromycin; each of the candidate RNA molecules further includes a pause sequence; the population of candidate RNA molecules includes at least 10^{13} different RNA molecules; the in vitro translation reaction is carried out in a lysate prepared from a eukaryotic cell or portion thereof (and is, for example, carried out in a reticulocyte lysate or wheat germ lysate); the selection step involves binding of the desired protein to an immobilized binding partner; the selection step involves assaying for a functional activity of the desired protein; the DNA molecule is amplified; and the method further involves transcribing an RNA molecule from the DNA molecule and repeating steps (a) through (d).

In other related aspects, the invention features an RNA-protein fusion selected by any of the methods of the invention; a ribonucleic acid covalently bonded though an amide bond to an amino acid sequence, the amino acid sequence being encoded by the

ribonucleic acid; and a ribonucleic acid which includes a translation initiation sequence and a start codon operably linked to a candidate protein coding sequence, the ribonucleic acid being covalently bonded to a peptide acceptor (for example, puromycin) at the 3' end of the candidate protein coding sequence.

As used herein, by a "population" is meant more than one molecule (for example, more than one RNA, DNA, or RNA-protein fusion molecule). Because the methods of the invention facilitate selections which begin, if desired, with large numbers of candidate molecules, a "population" according to the invention preferably means more than 10^9 molecules, more preferably, more than 10^{11} molecules, and, most preferably, more than 10^{13} molecules.

By a "protein" is meant any two or more amino acids joined by one or more peptide bonds. "Protein" and "peptide" are used interchangeably in this application.

By a "translation initiation sequence" is meant any sequence which is capable of providing a functional ribosome entry site. In bacterial systems, this region is sometimes referred to as a Shine-Dalgarno sequence.

By a "start codon" is meant three bases which signal the beginning of a protein coding sequence. Generally, these bases are AUG (or ATG); however, any other base triplet capable of being utilized in this manner may be substituted.

By "covalently bonded" to a peptide acceptor is meant that the peptide acceptor is joined to a "protein coding sequence" either directly through a covalent bond or indirectly through another covalently bonded sequence (for example, DNA corresponding to a pause site).

By a "peptide acceptor" is meant any molecule capable of being added to the C-terminus of a growing protein chain by the catalytic activity of the ribosomal peptidyl transferase function. Typically, such molecules contain (i) a nucleotide or nucleotide-like moiety (for example, adenosine or an adenosine analog (di-methylation at the N-6 amino position is acceptable)), (ii) an amino acid or amino acid-like moiety (for example, any of the 20 D- or L-amino acids or any amino acid analog thereof (for example, O-methyl

tyrosine or any of the analogs described by Ellman et al., Meth. Enzymol. **202**:301, 1991), and (iii) a linkage between the two (for example, an ester, amide, or ketone linkage at the 3' position or, less preferably, the 2' position); preferably, this linkage does not significantly perturb the pucker of the ring from the natural ribonucleotide conformation. Peptide acceptors may also possess a nucleophile, which may be, without limitation, an amino group, a hydroxyl group, or a sulfhydryl group.

By a peptide acceptor being positioned "at the 3' end" of a protein coding sequence is meant that the peptide acceptor molecule is either covalently bonded directly to the 3' end of the protein coding sequence or is covalently bonded indirectly to the 3' end of the protein coding sequence through another covalently bonded sequence (for example, DNA corresponding to a pause site).

By an "altered function" is meant any qualitative or quantitative change in the function of a molecule.

By a "pause sequence" is meant a nucleic acid sequence which causes a ribosome to slow its rate of translation.

The presently claimed invention provides a number of significant advantages. To begin with, it is the first example of this type of scheme for the selection and amplification of proteins. This technique overcomes the impasse created by the need to recover nucleotide sequences corresponding to desired, isolated proteins (since only nucleic acids can be replicated). In particular, many prior methods that allowed the isolation of proteins from partially or fully randomized pools did so through an <u>in vivo</u> step. Methods of this sort include monoclonal antibody technology (Schultz et al., J. Chem. Engng. News **68**:26 (1990)), phage display (McCafferty et al., Nature **348**:552 (1990)), peptide-lac repressor fusions (Cull et al., Proc. Natl. Acad. Sci. (USA) **89**:1865 (1992)), and classical genetic selections. Unlike the present technique, each of these methods relies on a topological link between the protein and the nucleic acid so that the information of the protein is retained and can be recovered in readable, nucleic acid form.

In addition, the present invention provides advantages over the stalled translation method (Tuerk and Gold, Science **249**:505 (1990); Irvine et al., J. Mol. Biol **222**:739 (1991); Korman et al., Proc. Natl. Acad. Sci. USA **79**:1844-1848 (1982); and Mattheakis et al., Proc. Natl. Acad. Sci. USA **91**:9022-9026 (1994)), a technique in which selection is for some property of a nascent protein chain that is still complexed with the ribosome and its mRNA. Unlike the stalled translation technique, the present method does not rely on maintaining the integrity of an mRNA: ribosome: nascent chain ternary complex, a complex that is very fragile and is therefore limiting with respect to the types of selections which are technically feasible.

The present method also provides advantages over the branched synthesis approach proposed by Brenner and Lerner (Proc. Natl. Acad. Sci. (USA) **89**:5381-5383 (1992)), in which DNA-peptide fusions are generated, and genetic information is theoretically recovered following one round of selection. Unlike the branched synthesis approach, the present method does not require the regeneration of a peptide from the DNA portion of a fusion (which, in the branched synthesis approach, is generally accomplished by individual rounds of chemical synthesis). This allows for repeated rounds of selection using populations of candidate molecules. In addition, unlike the branched synthesis technique, which is generally limited to the selection of fairly short sequences, the present method is applicable to the selection of protein molecules of considerable length.

In yet another advantage, the present selection technique can make use of very large and complex libraries of candidate sequences. In contrast, existing protein selection methods which rely on an in vivo step are typically limited to relatively small libraries of somewhat limited complexity. This advantage is particularly important when selecting functional protein sequences considering, for example, that 10^{13} possible sequences exist for a peptide of only 10 amino acids in length. In classical genetic techniques, lac repressor fusion approaches, and phage display methods, maximum complexities generally fall orders of magnitude below 10^{13} members.

The present technique also differs from prior approaches in that the selection step is context-independent. In many other selection schemes, the context in which, for example, an expressed protein is present can profoundly influence the nature of the library generated. For example, an expressed protein may not be properly expressed in a particular system or may not be properly displayed in vivo (for example, on the surface of a phage particle). Alternatively, the expression of a protein may actually interfere with one or more critical steps in a selection cycle, e.g., phage viability or infectivity, or lac repressor binding. These problems can result in the loss of functional molecules or in limitations on the nature of the selection procedures that may be applied.

Finally, the present method is advantageous because it provides control over the repertoire of proteins that may be tested. In certain techniques (for example, antibody selection), there exists little or no control over the nature of the starting pool. In yet other techniques (for example, lac fusions and phage display), the candidate pool must be expressed in the context of a fusion protein. In contrast, RNA-protein fusion constructs provide control over the nature of the candidate pools available for screening. In addition, the candidate pool size has the potential to be as high as RNA or DNA pools ($\sim 10^{15}$ members), limited only by the size of the in vitro translation reaction performed. And the makeup of the candidate pool depends completely on experimental design; random regions may be screened in isolation or within the context of a desired fusion protein, and most if not all possible sequences may be expressed in candidate pools of RNA-protein fusions.

Other features and advantages of the invention will be apparent from the following detailed description, and from the claims.

Detailed Description

The drawings will first briefly be described.

Brief Description of the Drawings

FIGURES 1A-C are schematic representations of steps involved in the production of RNA-protein fusions. Figure 1A illustrates a sample DNA construct for generation of an RNA portion of a fusion. Figure 1B illustrates the generation of an RNA/puromycin conjugate. And Figure 1C illustrates the generation of an RNA-protein fusion.

FIGURE 2 is a schematic representation of a generalized selection protocol according to the invention.

FIGURE 3 is a schematic representation of a synthesis protocol for minimal translation templates containing 3' puromycin. Step (A) shows the addition of protective groups to the reactive functional groups on puromycin (5'-OH and NH_2); as modified, these groups are suitably protected for use in phosphoramidite based oligonucleotide synthesis. The protected puromycin was attached to aminohexyl controlled pore glass (CPG) through the 2'OH group using the standard protocol for attachment of DNA through its 3'OH (Gait, Oligonucleotide Synthesis, A Practical Approach, The Practical Approach Series (IRL Press, Oxford, 1984)). In step (B), a minimal translation template (termed "43-P"), which contained 43 nucleotides, was synthesized using standard RNA and DNA chemistry (Millipore, Bedford, MA), deprotected using NH_4OH and TBAF, and gel purified. The template contained 13 bases of RNA at the 5' end followed by 29 bases of DNA attached to the 3' puromycin at its 5' OH. The RNA sequence contained (i) a Shine-Dalgarno consensus sequence complementary to five bases of 16S rRNA (Stormo et al., Nucleic Acids Research **10**:2971-2996 (1982); Shine and Dalgarno, Proc. Natl. Acad. Sci. USA **71**:1342-1346 (1974); and Steitz and Jakes, Proc. Natl. Acad. Sci. USA **72**:4734-4738 (1975)), (ii) a five base spacer, and (iii) a single AUG start codon. The DNA sequence was $dA_{27}dCdCP$, where "P" is puromycin.

FIGURE 4 is a schematic representation of a preferred method for the preparation of protected CPG-linked puromycin.

FIGURE 5 is a schematic representation showing possible modes of

methionine incorporation into a template of the invention. As shown in reaction (A), the template binds the ribosome, allowing formation of the 70S initiation complex. Fmet tRNA binds to the P site and is base paired to the template. The puromycin at the 3' end of the template enters the A site in an intramolecular fashion and forms an amide linkage to N-formyl methionine via the peptidyl transferase center, thereby deacylating the tRNA. Phenol/chloroform extraction of the reaction yields the template with methionine covalently attached. Shown in reaction (B) is an undesired intermolecular reaction of the template with puromycin containing oligonucleotides. As before, the minimal template stimulates formation of the 70S ribosome containing fmet tRNA bound to the P site. This is followed by entry of a second template in trans to give a covalently attached methionine.

FIGURES 6A-H are photographs showing the incorporation of ^{35}S methionine (^{35}S met) into translation templates. Figure 6A demonstrates magnesium (Mg^{+2}) dependence of the reaction. Figure 6B demonstrates base stability of the product; the change in mobility shown in this figure corresponds to a loss of the 5' RNA sequence of 43-P to produce the DNA-puromycin portion, termed 30-P. The retention of the label following base treatment was consistent with the formation of a peptide bond between ^{35}S methionine and the 3' puromycin of the template. Figure 6C demonstrates the inhibition of product formation in the presence of peptidyl transferase inhibitors. Figure 6C demonstrates the dependence of ^{35}S methionine incorporation on a template coding sequence. Figure 6E demonstrates DNA template length dependence of ^{35}S methionine incorporation. Figure 6F illustrates cis versus trans product formation using templates 43-P and 25-P. Figure 6G illustrates cis versus trans product formation using templates 43-P and 13-P. Figure 6H illustrates cis versus trans product formation using templates 43-P and 30-P in a reticulocyte lysate system.

FIGURES 7A-C are schematic illustrations of constructs for testing peptide fusion formation and selection. Figure 7A shows LP77 (“ligated-product,” “77" nucleotides long) (SEQ ID NO: 1). This sequence contains the c-myc monoclonal

antibody epitope tag EQKLISEEDL (SEQ ID NO: 2) flanked by a 5' start codon and a 3' linker. The 5' region contains a bacterial Shine-Dalgarno sequence identical to that of 43-P. The coding sequence was optimized for translation in bacterial systems. Figure 7B shows LP155 (ligated product, 155 nucleotides long) (SEQ ID NO: 3). This sequence contains the eukaryotic optimized code for generation of the peptide used to isolate the c-myc antibody. The 5' end contains a truncated version of the TMV upstream sequence (designated "TE). Figure 7C shows Pool #1 (SEQ ID NO: 4), an exemplary sequence to be used for peptide selection. The final seven amino acids from the original myc peptide were included in the template to serve as the 3' constant region required for PCR amplification of the template. This sequence is known not to be part of the antibody binding epitope.

FIGURE 8 is a photograph demonstrating the synthesis of RNA-protein fusions using templates 43-P, LP77, and LP155, and reticulocyte ("Retic") and wheat germ ("Wheat") translation systems. The left half of the figure illustrates ^{35}S methionine incorporation in each of the three templates. The right half of the figure illustrates the resulting products after RNase A treatment of each of the three templates to remove the RNA coding region; shown are ^{35}S methionine-labeled DNA-protein fusions. The DNA portion of each was identical to the oligo 30-P. Thus, differences in mobility were proportional to the length of the coding regions, consistent with the existence of proteins of different length in each case.

FIGURE 9 is a photograph demonstrating protease sensitivity of an RNA-protein fusion synthesized from LP155 and analyzed by denaturing polyacrylamide gel electrophoresis. Lane 1 contains ^{32}P labeled 30-P. Lanes 2-4, 5-7, and 8-10 contain the ^{35}S labeled translation templates recovered from reticulocyte lysate reactions either without treatment, with RNase A treatment, or with RNase A and proteinase K treatment, respectively.

FIGURE 10 is a photograph showing the results of immunoprecipitation reactions using in vitro translated 33 amino acid myc-epitope protein. Lanes 1 and 2

show the translation products of the myc epitope protein and β-globin templates, respectively. Lanes 3-5 show the results of immunoprecipitation of the myc-epitope peptide using a c-myc monoclonal antibody and PBS, DB, and PBSTDS wash buffers, respectively. Lanes 6-8 show the same immunoprecipitation reactions, but using the β-globin translation product.

FIGURE 11 is a photograph demonstrating immunoprecipitation of an RNA-protein fusion from an in vitro translation reaction. The picomoles of template used in the reaction are indicated. Lanes 1-4 show RNA142 (the RNA portion of fusion LP155), and lanes 5-7 show RNA-protein fusion LP155. After immunoprecipitation using a c-myc monoclonal antibody and protein G sepharose, the samples were treated with RNase A and T4 polynucleotide kinase, then loaded on a denaturing polyacrylamide gel to visualize the fusion. In lanes 1-4, no fusion was seen. In lanes 5-7, bands corresponding to the fusion were clearly visualized. The position of ^{32}P labeled 30-P is indicated.

FIGURE 12 is a graph showing a quantitation of fusion material obtained from an in vitro translation reaction. The intensity of the fusion bands shown in lanes 5-7 of Figure 11 and the 30-P band (isolated in a parallel fashion on dT_{25}, not shown) were quantitated on phosphorimager plates and plotted as a function of input LP155 concentration. From this analysis, it was calculated that $\sim 10^{12}$ fusions were formed per ml of translation reaction sample.

FIGURE 13 is a schematic representation of thiopropyl sepharose and dT_{25} agarose, and the ability of these substrates to interact with the RNA-protein fusions of the invention.

FIGURE 14 is a photograph showing the results of sequential isolation of fusions of the invention. Lane 1 contains ^{32}P labeled 30-P. Lanes 2 and 3 show LP155 isolated from translation reactions and treated with RNase A. In lane 2, LP155 was isolated sequentially, using thiopropyl sepharose followed by dT_{25} agarose. Lane 3 shows isolation using only dT_{25} agarose. The results indicated that the product contained a free thiol, likely the penultimate cysteine in the myc epitope coding sequence.

Described herein is a general method for the in vitro selection of proteins with desired functions using fusions in which these proteins are covalently linked to their own messenger RNAs. These RNA-protein fusions are synthesized by in vitro translation of mRNA pools containing a peptide acceptor attached to their 3' ends (Figure 1B). In one preferred embodiment, after readthrough of the open reading frame of the message, the ribosome pauses when it reaches the designed pause site, and the acceptor moiety occupies the ribosomal A site and accepts the nascent peptide chain from the peptidyl-tRNA in the P site to generate the RNA-protein fusion (Figure 1C). The covalent link between the protein and the RNA (in the form of an amide bond between the 3' end of the mRNA and the C-terminus of the protein which it encodes) allows the genetic information in the protein to be recovered and amplified (e.g., by PCR) following selection by reverse transcription of the RNA. Once the fusion is generated, screening is carried out based on the properties of the mRNA-protein fusion, or, alternatively, a cDNA may be generated using the mRNA template while it is attached to the protein to avoid any effect of the single-stranded RNA on the selection. When the mRNA-protein construct is used, selected fusions may be tested to determine which moiety (the protein, the RNA, or both) provides the desired function.

In one preferred embodiment, puromycin (which resembles tyrosyl adenosine) acts as the acceptor to attach the growing peptide to its mRNA. Puromycin is an antibiotic that acts by terminating peptide elongation. As a mimetic of aminoacyl-tRNA, it acts as a universal inhibitor of protein synthesis by binding the A site, accepting the growing peptide chain, and falling off the ribosome (at a Kd = 10^{-4} M) (Traut and Monro, J. Mol. Biol. **10**:63 (1964); Smith et al., J. Mol. Biol. **13**:617 (1965)). One of the most attractive features of puromycin is the fact that it forms a stable amide bond to the growing peptide chain, thus allowing for more stable fusions than potential acceptors that form unstable ester linkages. Other possible choices for acceptors include tRNA-like structures at the 3' end of the mRNA, as well as other compounds that act in a

manner similar to puromycin. Such compounds include, without limitation, any compound which possesses an amino acid linked to an adenine or an adenine-like compound, such as the amino acid nucleotides, phenylalanyl-adenosine (A-Phe), tyrosyl adenosine (A-Tyr), and alanyl adenosine (A-Ala), as well as amide-linked structures, such as phenylalanyl 3' deoxy 3' amino adenosine, alanyl 3' deoxy 3' amino adenosine, and tyrosyl 3' deoxy 3' amino adenosine; in any of these compounds, any of the naturally-occurring L-amino acids or their analogs may be utilized. In addition, a combined tRNA-like 3' structure-puromycin conjugate may also be used in the invention.

Shown in Figure 2 is a preferred selection scheme according to the invention. The steps involved in this selection are generally carried out as follows.

Step 1. Preparation of the DNA template. As a step toward generating the RNA-protein fusions of the invention, the RNA portion of the fusion is synthesized. This may be accomplished by direct chemical RNA synthesis or, more commonly, is accomplished by transcribing an appropriate double-stranded DNA template.

Such DNA templates may be created by any standard technique (including any technique of recombinant DNA technology, chemical synthesis, or both). In principle, any method that allows production of one or more templates containing a known, random, randomized, or mutagenized sequence may be used for this purpose. In one particular approach, an oligonucleotide (for example, containing random bases) is synthesized and is amplified (for example, by PCR) prior to transcription. Chemical synthesis may also be used to produce a random cassette which is then inserted into the middle of a known protein coding sequence (see, for example, chapter 8.2, Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons and Greene Publishing Company, 1994). This latter approach produces a high density of mutations around a specific site of interest in the protein.

An alternative to total randomization of a DNA template sequence is partial randomization, and a pool synthesized in this way is generally referred to as a "doped" pool. An example of this technique, performed on an RNA sequence, is described, for

example, by Ekland et al. (Nucl. Acids Research **23**:3231(1995)). Partial randomization may be performed chemically by biasing the synthesis reactions such that each base addition reaction mixture contains an excess of one base and small amounts of each of the others; by careful control of the base concentrations, a desired mutation frequency may be achieved by this approach. Partially randomized pools may also be generated using error prone PCR techniques, for example, as described in Beaudry and Joyce (Science **257**:635 (1992)) and Bartel and Szostak (Science **261**:1411 (1993)).

Numerous methods are also available for generating a DNA construct beginning with a known sequence and then creating a mutagenized DNA pool. Examples of such techniques are described in Ausubel et al. (supra, chapter 8) and Sambrook et al. (Molecular Cloning: A Laboratory Manual, chapter 15, Cold Spring Harbor Press, New York, 2nd ed. (1989)). Random sequences may also be generated by the "shuffling" technique outlined in Stemmer (Nature 370: **389** (1994)).

To optimize a selection scheme of the invention, the sequences and structures at the 5' and 3' ends of a template may also be altered. Preferably, this is carried out in two separate selections, each involving the insertion of random domains into the template proximal to the appropriate end, followed by selection. These selections may serve (i) to maximize the amount of fusion made (and thus to maximize the complexity of a library) or (ii) to provide optimized translation sequences. Further, the method may be generally applicable, combined with mutagenic PCR, to the optimization of translation templates both in the coding and non-coding regions.

Step 2. Generation of RNA. As noted above, the RNA portion of an RNA-protein fusion may be chemically synthesized using standard techniques of oligonucleotide synthesis. Alternatively, and particularly if longer RNA sequences are utilized, the RNA portion is generated by in vitro transcription of a DNA template. In one preferred approach, T7 polymerase is used to enzymatically generate the RNA strand. Other appropriate RNA polymerases for this use include, without limitation, the SP6, T3 and E. coli RNA polymerases (described, for example, in Ausubel et al. (supra, chapter

3).

Step 3. Ligation of Puromycin to the Template. Next, puromycin (or any other appropriate peptide acceptor) is covalently bonded to the template sequence. This step may be accomplished using T4 RNA ligase to attach the puromycin directly to the RNA sequence, or preferably the puromycin may be attached by way of a DNA "splint" using T4 DNA ligase or any other enzyme which is capable of joining together two nucleotide sequences (see Figure 1B) (see also, for example, Ausubel et al., supra, chapter 3, sections 14 and 15). tRNA synthetases may also be used to attach puromycin-like compounds to RNA. For example, phenylalanyl tRNA synthetase links phenylalanine to phenylalanyl-tRNA molecules containing a 3' amino group, generating RNA molecules with puromycin-like 3' ends (Fraser and Rich, Proc. Natl. Acad. Sci. USA **70**:2671 (1973)). Other peptide acceptors which may be used include, without limitation, any compound which possesses an amino acid linked to an adenine or an adenine-like compound, such as the amino acid nucleotides, phenylalanyl-adenosine (A-Phe), tyrosyl adenosine (A-Tyr), and alanyl adenosine (A-Ala), as well as amide-linked structures, such as phenylalanyl 3' deoxy 3' amino adenosine, alanyl 3' deoxy 3' amino adenosine, and tyrosyl 3' deoxy 3' amino adenosine; in any of these compounds, any of the naturally-occurring L-amino acids or their analogs may be utilized. A number of peptide acceptors are described, for example, in Krayevsky and Kukhanova, Progress in Nucleic Acids Research and Molecular Biology **23**:1 (1979).

Step 4. Generation and Recovery of RNA-Protein Fusions. To generate RNA-protein fusions, any in vitro translation system may be utilized. As shown below, eukaryotic systems are preferred, and two particularly preferred systems include the wheat germ and reticulocyte lysate systems. In principle, however, any translation system which allows formation of an RNA-protein fusion and which does not degrade the RNA portion of the fusion is useful in the invention. Examples of other useful eukaryotic systems include, without limitation, lysates from yeast, ascites, tumor cells (Leibowitz et al., Meth. Enzymol. **194**:536 (1991)) and xenopus oocyte eggs. Useful in vitro

translation systems from bacterial systems include, without limitation, those described in Zubay (Ann. Rev. Genet. **7**:267 (1973)); Chen and Zubay (Meth. Enzymol. **101**:44 (1983)); and Ellman (Meth. Enzymol. **202**:301 (1991)).

Once generated, RNA-protein fusions may be recovered from the in vitro translation reaction mixture by any standard technique of protein or RNA purification. Typically, protein purification techniques are utilized. As shown below, for example, purification of a fusion may be facilitated by the use of suitable chromatographic reagents such as dT_{25} agarose or thiopropyl sepharose. Purification, however, may also or alternatively involve purification based upon the RNA portion of the fusion; techniques for such purification are described, for example in Ausubel et al. (supra, chapter 4).

Step 5. Selection of the Desired RNA-Protein Fusion. Selection of a desired RNA-protein fusion may be accomplished by any means available to selectively partition or isolate a desired fusion from a population of candidate fusions. Examples of isolation techniques include, without limitation, selective binding, for example, to a binding partner which is directly or indirectly immobilized on a column, bead, or other solid support; and immunoprecipitation using an antibody specific for the protein moiety of the fusion. The first of these techniques makes use of an immobilized selection motif which can consist of any type of molecule to which binding is possible. A list of possible selection motif molecules is presented in Figure 2. Selection may also be based upon the use of substrate molecules attached to an affinity label (for example, substrate-biotin) which react with a candidate molecule, or upon any other type of interaction with a fusion molecule. In addition, proteins may be selected based upon their catalytic activity in a manner analogous to that described by Bartel and Szostak for the isolation of RNA enzymes (supra); according to that particular technique, desired molecules are selected based upon their ability to link a target molecule to themselves, and the functional molecules are then isolated based upon the presence of that target. Selection schemes for isolating novel or improved catalytic proteins using this same approach or any other functional selection are enabled by the present invention.

Step 7. Generation of a DNA Copy of the RNA Sequence using Reverse Transcriptase. If desired, a DNA copy of a selected RNA fusion sequence is readily available by reverse transcribing that RNA sequence using any standard technique (for example, using Superscript reverse transcriptase). This step may be carried out prior to the selection step, or following that step. Alternatively, the reverse transcription process may be carried out prior to the isolation of the fusion from the in vitro translation mixture.

Next, the DNA template is amplified, either as a partial or full-length double-stranded sequence. Preferably, in this step, full-length DNA templates are generated, using appropriate oligonucleotides and PCR amplification.

These steps, and the reagents and techniques for carrying out these steps, are now described in detail using particular examples. These examples are provided for the purpose of illustrating the invention, and should not be construed as limiting.

GENERATION OF TEMPLATES FOR RNA-PROTEIN FUSIONS

As shown in Figures 1A and 2, the selection scheme of the present invention preferably makes use of double-stranded DNA templates which include a number of design elements. The first of these elements is a promoter to be used in conjunction with a desired RNA polymerase for mRNA synthesis. As shown in Figure 1A and described herein, the T7 promoter is preferred, although any promoter capable of directing synthesis from a linear double-stranded DNA may be used.

The second element of the template shown in Figure 1A is termed the 5' untranslated region (or 5'UTR) and corresponds to the RNA upstream of the translation start site. Shown in Figure 1A is a preferred 5'UTR (termed "TE") which is a deletion mutant of the Tobacco Mosaic Virus 5' untranslated region and, in particular, corresponds to the bases directly 5' of the TMV translation start; the sequence of this UTR is as follows: rGrGrG rArCrA rArUrU rArCrU rArUrU rUrArC rArArU rUrArC rA (with the first 3 G nucleotides being inserted to augment transcription) (SEQ ID NO: 5). Any other

appropriate 5' UTR may be utilized (see, for example, Kozak, Microbiol. Rev. **47**:1 (1983)).

The third element shown in Figure 1A is the translation start site. In general, this is an AUG codon. However, there are examples where codons other than AUG are utilized in naturally-occurring coding sequences, and these codons may also be used in the selection scheme of the invention.

The fourth element in Figure 1A is the open reading frame of the protein (termed ORF), which encodes the protein sequence. This open reading frame may encode any naturally-occurring, random, randomized, mutagenized, or totally synthetic protein sequence.

The fifth element shown in Figure 1A is the 3' constant region. This sequence facilitates PCR amplification of the pool sequences and ligation of the puromycin-containing oligonucleotide to the mRNA. If desired, this region may also include a pause site, a sequence which causes the ribosome to pause and thereby allows additional time for an acceptor moiety (for example, puromycin) to accept a nascent peptide chain from the peptidyl-tRNA; this pause site is discussed in more detail below.

To develop the present methodology, RNA-protein fusions were initially generated using highly simplified mRNA templates containing 1-2 codons. This approach was taken for two reasons. First, templates of this size could readily be made by chemical synthesis. And, second, a small open reading frame allowed critical features of the reaction, including efficiency of linkage, end heterogeneity, template dependence, and accuracy of translation, to be readily assayed.

Design of Construct. A basic construct was used for generating test RNA-protein fusions. The molecule consisted of a mRNA containing a Shine-Dalgarno (SD) sequence for translation initiation which contained a 3 base deletion of the SD sequence from ribosomal protein L1 and which was complementary to 5 bases of 16S rRNA (i.e., rGrGrA rGrGrA rCrGrA rA) (SEQ ID NO: 6) (Stormo et al., Nucleic Acids Research **10**:2971-2996 (1982); Shine and Dalgarno, Proc. Natl. Acad. Sci. USA **71**:1342-1346

(1974); and Steitz and Jakes, Proc. Natl. Acad. Sci. USA **72**:4734-4738 (1975)), (ii) an AUG start codon, (iii) a DNA linker to act as a pause site (i.e., 5'-$(dA)_{27}$), (iv) dCdC-3', and (v) a 3' puromycin (P). The poly dA sequence was chosen because it was known to template tRNA poorly in the A site (Morgan et al., J. Mol. Biol. **26**:477-497 (1967); Ricker and Kaji, Nucleic Acid Research **19**:6573-6578 (1991)) and was designed to act as a good pause site. The length of the oligo dA linker was chosen to span the ~60-70 Å distance between the decoding site and the peptidyl transfer center of the ribosome. The dCdCP mimicked the CCA end of a tRNA and was designed to facilitate binding of the puromycin to the A site of the ribosome.

Chemical Synthesis of Minimal Template 43-P. To synthesize construct 43-P (shown in Figure 3), puromycin was first attached to a solid support in such a way that it would be compatible with standard phosphoramidite oligonucleotide synthesis chemistry. The synthesis protocol for this oligo is outlined schematically in Figure 3 and is described in more detail below. To attach puromycin to a controlled pore glass (CPG) solid support, the amino group was protected with a trifluoroacetyl group as described in Applied Biosystems User Bulletin #49 for DNA synthesizer model 380 (1988). Next, protection of the 5' OH was carried out using a standard DMT-Cl approach (Gait, Oligonucleotide Synthesis a practical approachThe Practical Approach Series (IRL Press, Oxford, 1984)), and attachment to aminohexyl CPG through the 2' OH was effected in exactly the same fashion as the 3' OH would be used for attachment of a deoxynucleoside (see Fig. 3 and Gait, supra, p. 47). The 5' DMT-CPG-linked protected puromycin was then suitable for chain extension with phosphoramidite monomers. The synthesis of the oligo proceeded in the 3' -> 5' direction in the order: (i) 3' puromycin, (ii) pdCpdC, (iii) ~27 units of dA as a linker, (iv) AUG, and (v) the Shine-Dalgarno sequence. The sequence of the 43-P construct is shown below.

Synthesis of CPG Puromycin. The synthesis of protected CPG puromycin followed the general path used for deoxynucleosides as previously outlined (Gait, Oligonucleotide Synthesis, A Practical Approach, The Practical Approach Series (IRL

Press, Oxford, 1984)). Major departures included the selection of an appropriate N blocking group, attachment at the 2' OH to the solid support, and the linkage reaction to the solid support. In the case of the latter, the reaction was carried out at very low concentrations of activated nucleotide as this material was significantly more precious than the solid support. The resulting yield (~20 μmol/g support) was quite satisfactory considering the dilute reaction conditions.

Synthesis of N-Trifluoroacetyl Puromycin. 267 mg (0.490 mmol) Puromycin*HCl was first converted to the free base form by dissolving in water, adding pH 11 carbonate buffer, and extracting (3X) into chloroform. The organic phase was evaporated to dryness and weighed (242 mg, 0.513 mmol). The free base was then dissolved in 11 ml dry pyridine and 11 ml dry acetonitrile, and 139 μl (2.0 mmol) triethanolamine acetate (TEA) and 139 μl (1.0 mmol) of trifluoroacetic anhydride (TFAA) were added with stirring. TFAA was then added to the turbid solution in 20 μl aliquots until none of the starting material remained, as assayed by thin layer chromatography (tlc) (93:7, Chloroform/MeOH) (a total of 280 μl). The reaction was allowed to proceed for one hour. At this point, two bands were revealed by thin layer chromatography, both of higher mobility than the starting material. Workup of the reaction with NH_4OH and water reduced the product to a single band. Silica chromatography (93:7 Chloroform/MeOH) yielded 293 mg (0.515 mmol) of the product, N-TFA-Pur. The product of this reaction is shown schematically in Figure 4.

Synthesis of N-Trifluoroacetyl 5'-DMT Puromycin. The product from the above reaction was aliquoted and coevaporated 2X with dry pyridine to remove water. Multiple tubes were prepared to test multiple reaction conditions. In a small scale reaction, 27.4 mg (48.2 μmoles) N-TFA-Pur were dissolved in 480 μl of pyridine containing 0.05 eq of DMAP and 1.4 eq TEA. To this mixture, 20.6 mg of trityl chloride (60 μmol) was added, and the reaction was allowed to proceed to completion with stirring. The reaction was stopped by addition of an equal volume of water (approximately 500 μl) to the solution. Because this reaction appeared successful , a

large scale version was performed. In particular, 262 mg (0.467 mmol) N-TFA-Pur was dissolved in 2.4 ml pyridine followed by addition of 1.4 eq of TEA, 0.05 eq of DMAP, and 1.2 eq of trityl chloride. After approximately two hours, an additional 50 mg (0.3 eq) dimethoxytrityl*Cl (DMT*Cl) was added, and the reaction was allowed to proceed for 20 additional minutes. The reaction was stopped by the addition of 3 ml of water and coevaporated 3X with CH_3CN. The reaction was purified by 95:5 Chloroform/MeOH on a 100 ml silica (dry) 2 mm diameter column. Due to incomplete purification, a second identical column was run with 97.5:2.5 Chloroform/MeOH. The total yield was 325 mg or 0.373 mmol (or a yield of 72%). The product of this reaction is shown schematically in Figure 4.

Synthesis of N-Trifluoroacetyl, 5'-DMT, 2' Succinyl Puromycin. In a small scale reaction, 32 mg (37 μmol) of the product synthesized above was combined with 1.2 eq of DMAP dissolved in 350 μl of pyridine. To this solution, 1.2 equivalents of succinic anhydride was added in 44 μl of dry CH_3CN and allowed to stir overnight. Thin layer chromatography revealed little of the starting material remaining. In a large scale reaction, 292 mg (336 mmol) of the previous product was combined with 1.2 eq DMAP in 3 ml of pyridine. To this, 403 μl of 1M succinic anhydride in dry CH_3CN was added, and the mixture was allowed to stir overnight. Thin layer chromatography again revealed little of the starting material remaining. The two reactions were combined, and an additional 0.2 eq of DMAP and succinate were added. The product was coevaporated with toluene 1X and dried to a yellow foam in high vacuum. $MeCl_2$ was added (20 ml), and this solution was extracted twice with 15 ml of 10% ice cold citric acid and then twice with pure water. The product was dried, redissolved in 2 ml of $MeCl_2$, and precipitated by addition of 50 ml of hexane with stirring. The product was then vortexed and centrifuged at 600 rpm for 10 minutes in the clinical centrifuge. The majority of the eluent was drawn off, and the rest of the product was dried, first at low vacuum, then at high vacuum in a dessicator. The yield of this reaction was approximately 260 mmol for a stepwise yield of ~70 %.

Synthesis of N-Trifluoroacetyl 5'-DMT, 2' Succinyl, CPG Puromycin. The product from the previous step was next dissolved with 1 ml of dioxane followed by 0.2 ml dioxane/0.2 ml pyridine. To this solution, 40 mg of p-nitrophenol and 140 mg of dicyclohexylcarbodiimide (DCC) was added, and the reaction was allowed to proceed for 2 hours. The insoluble cyclohexyl urea produced by the reaction was removed by centrifugation, and the product solution was added to 5 g of CPG suspended in 22 ml of dry DMF and stirred overnight. The resin was then washed with DMF, methanol, and ether, and dried. The resulting resin was assayed as containing 22.6 mmol of trityl per g, well within the acceptable range for this type of support. The support was then capped by incubation with 15 ml of pyridine, 1 ml of acetic anhydride, and 60 mg of DMAP for 30 minutes. The resulting column material produced a negative (no color) ninhydrin test, in contrast to the results obtained before blocking in which the material produced a dark blue color reaction. The product of this reaction is shown schematically in Figure 4.

Synthesis of mRNA-Puromycin Conjugate. As discussed above, a puromycin tethered oligo may be used in either of two ways to generate a mRNA-puromycin conjugate which acts as a translation template. For extremely short open reading frames, the puromycin oligo is typically extended chemically with RNA or DNA monomers to create a totally synthetic template. When longer open reading frames are desired, the RNA or DNA oligo is generally ligated to the 3' end of an mRNA using a DNA splint and T4 DNA ligase as described by Moore and Sharp (Science **256**:992 (1992)).

IN VITRO TRANSLATION AND TESTING OF RNA-PROTEIN FUSIONS

The templates generated above were translated in vitro using both bacterial and eukaryotic in vitro translation systems as follows.

In Vitro Translation of Minimal Templates. 43-P and related RNA-puromycin conjugates were added to several different in vitro translation systems including: (i) the S30 system derived from E. coli MRE600 (Zubay, Ann. Rev. Genet. 7:267 (1973);

Collins, Gene **6**:29 (1979); Chen and Zubay, Methods Enzymol, **101**:44 (1983); Pratt, in Transcription and Translation: A Practical Approach, B. D. Hammes, S. J. Higgins, Eds. (IRL Press, Oxford, 1984) pp. 179-209; and Ellman et al., Methods Enzymol. **202**:301 (1991)) prepared as described by Ellman et. al. (Methods Enzymol. **202**:301 (1991)); (ii) the ribosomal fraction derived from the same strain, prepared as described by Kudlicki et al. (Anal. Chem. **206**:389 (1992)); and (iii) the S30 system derived from E. coli BL21, prepared as described by Lesley et al. (J. Biol. Chem. **266**:2632 (1991)). In each case, the premix used was that of Lesley et al. (J. Biol. Chem. **266**:2632 (1991)), and the incubations were 30 minutes in duration.

Testing the Nature of the Fusion. The 43-P template was first tested using S30 translation extracts from E. coli. Figure 5 (Reaction "A") demonstrates the desired intramolecular (cis) reaction wherein 43-P binds the ribosome and acts as a template for and an acceptor of fMet at the same time. The incorporation of ^{35}S-methionine and its position in the template was first tested, and the results are shown in Figures 6A and 6B. After extraction of the in vitro translation reaction mixture with phenol/chloroform and analysis of the products by SDS-PAGE, an ^{35}S labeled band appeared with the same mobility as the 43-P template. The amount of this material synthesized was dependent upon the Mg^{2+} concentration (Figure 6A). The optimum Mg^{2+} concentration appeared to be between 9 and 18 mM, which was similar to the optimum for translation in this system (Zubay, Ann. Rev. Genet. 7:267 (1973); Collins, Gene **6**:29 (1979); Chen and Zubay, Methods Enzymol, **101**:44 (1983); Pratt, in Transcription and Translation: A Practical Approach, B. D. Hammes, S. J. Higgins, Eds. (IRL Press, Oxford, 1984) pp. 179-209; Ellman et al., Methods Enzymol. **202**:301 (1991); Kudlicki et al., Anal. Chem. **206**:389 (1992); and Lesley et al., J. Biol. Chem. **266**:2632 (1991)). Furthermore, the incorporated label was stable to treatment with NH_4OH (Figure 6B), indicating that the label was located on the 3' half of the molecule (the base-stable DNA portion) and was attached by a base-stable linkage, as expected for an amide bond between puromycin and fMet.

Ribosome and Template Dependence. To demonstrate that the reaction observed above occurred on the ribosome, the effects of specific inhibitors of the peptidyl transferase function of the ribosome were tested (Figure 6C), and the effect of changing the sequence coding for methionine was examined (Figure 6D). Figure 6C demonstrates clearly that the reaction was strongly inhibited by the peptidyl transferase inhibitors, virginiamycin, gougerotin, and chloramphenicol (Monro and Vazquez, J. Mol. Biol. **28**:161-165 (1967); and Vazquez and Monro, Biochemica et Biophysical Acta **142**:155-173 (1967)). Figure 6D demonstrates that changing a single base in the template from A to C abolished incorporation of ^{35}S methionine at 9 mM Mg^{2+}, and greatly decreased it at 18 mM (consistent with the fact that high levels of Mg^{2+} allow misreading of the message). These experiments demonstrated that the reaction occurred on the ribosome in a template dependent fashion.

Linker Length. Also tested was the dependence of the reaction on the length of the linker (Figure 6E). The original template was designed so that the linker spanned the distance from the decoding site (occupied by the AUG of the template) to the acceptor site (occupied by the puromycin moiety), a distance which was approximately the same length as the distance between the anticodon loop and the acceptor stem in a tRNA, or about 60-70 Å. The first linker tested was 30 nucleotides in length, based upon a minimum of 3.4 Å per base (≥ 102 Å). In the range between 30 and 21 nucleotides (n = 27 - 18; length ≥ 102 - 71 Å), little change was seen in the efficiency of the reaction.

Intramolecular vs. Intermolecular Reactions. Finally, we tested whether the reaction occurred in an intramolecular fashion (Figure 5, Reaction "A") as desired or intermolecularly (Figure 5, Reaction "B"). This was tested by adding oligonucleotides with 3' puromycin but no ribosome binding sequence (i.e., templates 25-P, 13-P, and 30-P) to the translation reactions containing the 43-P template (Figures 6F, 6G, and 6H). If the reaction occurred by an intermolecular mechanism, the shorter oligos would also be labeled. As demonstrated in Figures 6F-H, there was little incorporation of ^{35}S methionine in the three shorter oligos, indicating that the reaction occurred primarily in

an intramolecular fashion. The sequences of 25-P, 13-P, and 30-P are shown below.

Reticulocyte Lysate. Figure 6H demonstrates that ^{35}S-methionine may be incorporated in the 43-P template using a rabbit reticulocyte lysate (see below) for in vitro translation, in addition to the E. coli lysates used above. This reaction occurred primarily in an intramolecular mechanism, as desired.

SYNTHESIS AND TESTING OF FUSIONS CONTAINING A C-MYC EPITOPE TAG

Exemplary fusions were also generated which contained, within the protein portion, the epitope tag for the c-myc monoclonal antibody 9E10 (Evan et al., Mol. Cell Biol. **5**:3610 (1985)).

Design of Templates. Three initial epitope tag templates (i.e., LP77, LP155, and Pool #1) were designed and are shown in Figures 7A-C. The first two templates contained the c-myc epitope tag sequence EQKLISEEDL (SEQ ID NO: 2), and the third template was the design used in the synthesis of a random selection pool. LP77 encoded a 12 amino acid sequence, with the codons optimized for bacterial translation. LP155 and its derivatives contained a 33 amino acid mRNA sequence in which the codons were optimized for eukaryotic translation. The encoded amino acid sequence of MAEEQKLISEEDLLRKRREQKLKHKLEQLRNSCA (SEQ ID NO: 7) corresponded to the original peptide used to isolate the 9E10 antibody. Pool#1 contained 27 codons of NNG/C (to generate random peptides) followed by a sequence corresponding to the last seven amino acids of the myc peptide (which were not part of the myc epitope sequence). These sequences are shown below.

Reticulocyte vs. Wheat Germ In Vitro Translation Systems. The 43-P, LP77, and LP155 templates were tested in both rabbit reticulocyte and wheat germ extract (Promega, Boehringer Mannheim) translation systems (Figure 8). Translations were performed at 30°C for 60 minutes. Templates were isolated using dT_{25} agarose at 4°C. Templates were eluted from the agarose using 15 mM NaOH, 1mM EDTA, neutralized

with NaOAc/HOAc buffer, immediately ethanol precipitated (2.5 - 3 vol), washed (with 100% ethanol), and dried on a speedvac concentrator. Figure 8 shows that ^{35}S methionine was incorporated into all three templates, in both the wheat germ and reticulocyte systems. Less degradation of the template was observed in the fusion reactions from the reticulocyte system and, accordingly, this system is preferred for the generation of RNA-protein fusions. In addition, in general, eukaryotic systems are preferred over bacterial systems. Because eukaryotic cells tend to contain lower levels of nucleases, mRNA lifetimes are generally 10-100 times longer in these cells than in bacterial cells. In experiments using one particular E. coli translation system, generation of fusions was not observed using a template encoding the c-myc epitope; labeling the template in various places demonstrated that this was likely due to degradation of the RNA portion of the template.

To examine the peptide portion of these fusions, samples were treated with RNase to remove the coding sequence. Following this treatment, the 43-P product ran with almost identical mobility to the ^{32}P labeled 30-P oligo, consistent with a very small peptide (perhaps only methionine) added to 30-P. For LP77, removal of the coding sequence produced a product with lower mobility than the 30-P oligo, consistent with the notion that a 12 amino acid peptide was added to the puromycin. Finally, for LP155, removal of the coding sequence produced a product of yet lower mobility, consistent with a 33 amino acid sequence attached to the 30-P oligo. No oligo was seen in the RNase-treated LP155 reticulocyte lane due to a loading error. In Figure 9, the mobility of this product was shown to be the same as the product generated in the wheat germ extract. In sum, these results indicated that RNase resistant products were added to the ends of the 30-P oligos, that the sizes of the products were proportional to the length of the coding sequences, and that the products were quite homogeneous in size. In addition, although both systems produced similar fusion products, the reticulocyte system appeared superior due to higher template stability.

Sensitivity to RNase A and Proteinase K. In Figure 9, sensitivity to RNase A

and proteinase K were tested using the LP155 fusion. As shown in lanes 2-4, incorporation of ^{35}S methionine was demonstrated for the LP155 template. When this product was treated with RNase A, the mobility of the fusion decreased, but was still significantly higher than the ^{32}P labeled 30-P oligonucleotide, consistent with the addition of a 33 amino acid peptide to the 3' end. When this material was also treated with proteinase K, the ^{35}S signal completely disappeared, again consistent with the notion that the label was present in a peptide at the 3' end of the 30-P fragment.

Immunoprecipitation Experiments. In an experiment designed to illustrate the efficacy of immunoprecipitating an mRNA-peptide fusion, we attempted to immunoprecipitate a free c-myc peptide generated by in vitro translation. Figure 10 shows the results of these experiments assayed on an SDS PAGE peptide gel. Lanes 1 and 2 show the labeled material from translation reactions containing either RNA142 (the RNA portion of LP155) or β-globin mRNA. Lanes 3-8 show the immunoprecipitation of these reaction samples using the c-myc monoclonal antibody 9E10, under several different buffer conditions (described below). Lanes 3-5 show that the peptide derived from RNA142 was effectively immunoprecipitated, with the best case being lane 4 where ~83% of the total TCA precipitable counts were isolated. Lanes 6-8 show little of the β-globin protein, indicating a purification of >100 fold. These results indicated that the peptide coded for by RNA142 (and by LP155) can be quantitatively isolated by this immunoprecipitation protocol.

Immunoprecipitation of the Fusion. We next tested the ability to immunoprecipitate a chimeric RNA-peptide product, using an LP155 translation reaction and the c-myc monoclonal antibody 9E10 (Figure 11). The translation products from a reticulocyte reaction were isolated by immunoprecipitation (as described herein) and treated with 1 ug of RNase A at room temperature for 30 minutes to remove the coding sequence. This generated a 5'OH, which was ^{32}P labeled with T4 polynucleotide kinase and assayed by denaturing PAGE. Figure 11 demonstrates that a product with a mobility similar to that seen for the fusion of the c-myc epitope with 30-P generated by RNase

treatment of the LP155 fusion (see above) was isolated. In Figure 12, the quantity of fusion protein isolated was determined and was plotted against the amount of unmodified 30-P (not shown in this figure). While less than 1% of the input template was modified, the results still indicated that approximately 10^{12} molecules of fusion could be generated per ml of in vitro translation reaction mix.

Sequential Isolation. As a further confirmation of the nature of the in vitro translated LP155 template product, we examined the behavior of this product on two different types of chromatography media. Thiopropyl (TP) sepharose allows the isolation of a product containing a free cysteine (for example, the LP155 product which has a cysteine residue adjacent to the C terminus) (Figure 13). Similarly, dT_{25} agarose allows the isolation of templates containing a poly dA sequence (for example, 30-P) (Figure 13). Figure 14 demonstrates that sequential isolation on TP sepharose followed by dT_{25} agarose produced the same product as isolation on dT_{25} agarose alone. The fact that the in vitro translation product contained both a poly-A tract and a free thiol strongly indicated that the translation product was the desired RNA-peptide fusion.

The above results are consistent with the ability to synthesize mRNA-peptide fusions and to recover them intact from in vitro translation extracts. The peptide portions of fusions so synthesized appeared to have the intended sequences as demonstrated by immunoprecipitation and isolation using appropriate chromatographic techniques. According to the results presented above, the reactions are intramolecular and occur in a template dependent fashion. Finally, even with a template modification of less than 1%, the present system facilitates selections based on candidate complexities of about 10^{13} molecules.

C-Myc Epitope Recovery Selection. To select additional c-myc epitopes, a large library of translation templates (for example, 10^{15} members) is generated containing a randomized region (see Figure 7C and below). This library is used to generate $\sim 10^{12}$ - 10^{13} fusions (as described herein) which are treated with the anti-c-myc antibody (for example, by immunoprecipitation or using an antibody immobilized on a column or other

solid support) to enrich for c-myc-encoding templates in repeated rounds of in vitro selection.

DETAILED MATERIALS AND METHODS

Described below are detailed materials and methods used in the examples presented above.

Sequences. A number of oligonucleotides were used above for the generation of RNA-protein fusions. These oligonucleotides have the following sequences.

NAME SEQUENCE

30-P 5'AAA AAA AAA AAA AAA AAA AAA AAA AAA CCP (SEQ ID NO:8)

13-P 5'AAA AAA AAA ACC P (SEQ ID NO: 9)

25-P 5'CGC GGT TTT TAT TTT TTT TTT TCC P (SEQ ID NO: 10)

43-P 5'rGrGrA rGrGrA rCrGrA rArArU rGAA AAA AAA AAA AAA AAA AAA AAA AAA ACC P (SEQ ID NO: 11)

43-P [CUG] 5'rGrGrA rGrGrA rCrGrA rArCrU rGAA AAA AAA AAA AAA AAA AAA AAA AAA ACC P (SEQ ID NO: 12)

40-P 5'rGrGrA rGrGrA rCrGrA rArCrU rGAA AAA AAA AAA AAA AAA AAA AAA ACC P (SEQ ID NO: 13)

37-P 5'rGrGrA rGrGrA rCrGrA rArCrU rGAA AAA AAA AAA AAA AAA AAA ACC P (SEQ ID NO: 14)

34-P 5'rGrGrA rGrGrA rCrGrA rArCrU rGAA AAA AAA AAA AAA AAA ACC

P (SEQ ID NO: 15)

31-P 5'rGrGrA rGrGrA rCrGrA rArCrU rGAA AAA AAA AAA AAA ACC P (SEQ ID NO: 16)

LP77 5'rGrGrG rArGrG rArCrG rArArA rUrGrG rArArC rArGrA rArArC rUrGrA rUrCrU rCrUrG rArArG rArArG rArCrC rUrGrA rArC AAA AAA AAA AAA AAA AAA AAA AAA AAA CCP (SEQ ID NO: 1)

LP155 5'rGrGrG rArCrA rArUrU rArCrU rArUrU rUrArC rArArU rUrArC rA rArUrG rGrCrU rGrArA rGrArA rCrArG rArArA rCrUrG rArUrC rUrCrU rGrArA rGrArA rGrArC rCrUrG rCrUrG rCrGrU rArArA rCrGrU rCrGrU rGrArA rCrArG rCrUrG rArArA rCrArC rArArA rCrUrG rGrArA rCrArG rCrUrG rCrGrU rArArC rUrCrU rUrGrC rGrCrU AAA AAA AAA AAA AAA AAA AAA AAA AAA CCP (SEQ ID NO: 3)

LP160 5' 5'rGrGrG rArCrA rArUrU rArCrU rArUrU rUrArC rArArU rUrArC rA rArUrG rNrNrS rCrArG rCrUrG rCrGrU rArArC rUrCrU rUrGrC rGrCrU AAA AAA AAA AAA AAA AAA AAA AAA AAA CCP (SEQ ID NO: 17)

All oligonucleotides are listed in the 5' to 3' direction. Ribonucleotide bases are indicated by lower case "r" prior to the nucleotide designation; P is puromycin; rN indicates equal amounts of rA, rG, rC, and rU; rS indicates equal amounts of rG and rC; and all other base designations indicate DNA oligonucleotides.

Chemicals. Puromycin HCl, long chain alkylamine controlled pore glass,

gougerotin, chloramphenicol, virginiamycin, DMAP, dimethyltrityl chloride, and acetic anhydride were obtained from Sigma Chemical (St. Louis, MO). Pyridine, dimethylformamide, toluene, succinic anhydride, and para-nitrophenol were obtained from Fluka Chemical (Ronkonkoma, NY). Beta-globin mRNA was obtained from Novagen (Madison, WI). TMV RNA was obtained from Boehringer Mannheim (Indianapolis, IN).

Enzymes. Proteinase K was obtained from Promega (Madison, WI). DNase-free RNAase was either produced by the protocol of Sambrook et al. (supra) or purchased from Boehringer Mannheim. T7 polymerase was made by the published protocol of Grodberg and Dunn (J. Bacteriol. **170**:1245 (1988)) with the modifications of Zawadzki and Gross (Nucl. Acids Res. **19**:1948 (1991)). T4 DNA ligase was obtained from New England Biolabs (Beverly, MA).

Quantitation of Radiolabel Incorporation. For radioactive gels bands, the amount of radiolabel (^{35}S or ^{32}P) present in each band was determined by quantitation either on a Betagen 603 blot analyzer (Betagen, Waltham, MA) or using phosphorimager plates (Molecular Dynamics, Sunnyvale, CA). For liquid and solid samples, the amount of radiolabel (^{35}S or ^{32}P) present was determined by scintillation counting (Beckman, Columbia, MD).

Gel Images. Images of gels were obtained by autoradiography (using Kodak XAR film) or using phosphorimager plates (Molecular Dynamics).

Synthesis of CPG Puromycin. Detailed protocols for synthesis of CPG-puromycin are outlined above.

Enzymatic Reactions. In general, the preparation of nucleic acids for kinase, transcription, PCR, and translation reactions using E. coli extracts was the same. Each preparative protocol began with extraction using an equal volume of 1:1 phenol/chloroform, followed by centrifugation and isolation of the aqueous phase. Sodium acetate (pH 5.2) and spermidine were added to a final concentration of 300 mM and 1 mM respectively, and the sample was precipitated by addition of 3 volumes of

100% ethanol and incubation at -70°C for 20 minutes. Samples were centrifuged at >12,000 g, the supernatant was removed, and the pellets were washed with an excess of 95% ethanol, at 0°C. The resulting pellets were then dried under vacuum and resuspended.

Oligonucleotides. All synthetic DNA and RNA was synthesized on a Millipore Expedite synthesizer using standard chemistry for each as supplied from the manufacturer (Milligen, Bedford, MA). Oligonucleotides containing 3' puromycin were synthesized using CPG puromycin columns packed with 30-50 mg of solid support (~20 μmole puromycin/gram). Oligonucleotides containing a 3' biotin were synthesized using 1 μmole bioteg CPG columns from Glen Research (Sterling, VA). Oligonucleotides containing a 5' biotin were synthesized by addition of bioteg phosphoramidite (Glen Research) as the 5' base. Oligonucleotides to be ligated to the 3' ends of RNA molecules were either chemically phosphorylated at the 5' end (using chemical phosphorylation reagent from Glen Research) prior to deprotection or enzymatically phosphorylated using ATP and T4 polynucleotide kinase (New England Biolabs) after deprotection. Samples containing only DNA (and 3' puromycin or 3' biotin) were deprotected by addition of 25% NH_4OH followed by incubation for 12 hours at 55°C. Samples containing RNA monomers (e.g., 43-P) were deprotected by addition of ethanol (25% (v/v)) to the NH_4OH solution and incubation for 12 hours at 55°C. The 2'OH was deprotected using 1M TBAF in THF (Sigma) for 48 hours at room temperature. TBAF was removed using a NAP-25 Sephadex column (Pharmacia, Piscataway, NJ).

Deprotected DNA and RNA samples were then purified using denaturing PAGE, followed by either soaking or electro-eluting from the gel using an Elutrap (Schleicher and Schuell, Keene, NH) and desalting using either a NAP-25 Sephadex column or ethanol precipitation as described above.

Myc DNA construction. Two DNA templates containing the c-myc epitope tag were constructed. The first template was made from a combination of the oligonucleotides 64.27 (5'-GTT CAG GTC TTC TTG AGA GAT CAG TTT CTG TTC

CAT TTC GTC CTC CCT ATA GTG AGT CGT ATT A-3') (SEQ ID NO: 18) and 18.109 (5'-TAA TAC GAC TCA CTA TAG-3') (SEQ ID NO: 19). Transcription using this template produced RNA 47.1 which coded for the peptide MEQKLISEEDLN (SEQ ID NO: 20). Ligation of RNA 47.1 to 30-P yielded LP77 shown in Figure 7A.

The second template was made first as a single oligonucleotide 99 bases in length, having the designation RWR 99.6 and the sequence 5'AGC GCA AGA GTT ACG CAG CTG TTC CAG TTT GTG TTT CAG CTG TTC ACG ACG TTT ACG CAG CAG GTC TTC TTC AGA GAT CAG TTT CTG TTC TTC AGC CAT-3' (SEQ ID NO: 21). Double stranded transcription templates containing this sequence were constructed by PCR with the oligos RWR 21.103 (5'-AGC GCA AGA GTT ACG CAG CTG-3') (SEQ ID NO: 22) and RWR 63.26 (5'TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA ATG GCT GAA GAA CAG AAA CTG-3') (SEQ ID NO: 23) according to published protocols (Ausubel et al., <u>supra</u>, chapter 15). Transcription using this template produced an RNA referred to as RNA142 which coded for the peptide MAEEQKLISEEDLLRKRREQLKHKLEQLRNSCA (SEQ ID NO: 24). This peptide contained the sequence used to raise monoclonal antibody 9E10 when conjugated to a carrier protein (Oncogene Science Technical Bulletin). RNA142 was 125 nucleotides in length, and ligation of RNA142 to 30-P produced LP155 shown in Figure 7B.

<u>Randomized Pool Construction.</u> The randomized pool was constructed as a single oligonucleotide 130 bases in length denoted RWR130.1. Beginning at the 3' end, the sequence was 3' CCCTGTTAATGATAAATGTTAATGTTAC $(NNS)_{27}$ GTC GAC GCA TTG AGA TAC CGA-5' (SEQ ID NO: 25). N denotes a random position, and this sequence was generated according to the standard synthesizer protocol. S denotes an equal mix of dG and dC bases. PCR was performed with the oligonucleotides 42.108 (5'-TAA TAC GAC TCA CTA TAG GGA CAA TTA CTA TTT ACA ATT ACA) (SEQ ID NO: 26) and 21.103 (5'-AGC GCA AGA GTT ACG CAG CTG) (SEQ ID NO: 27). Transcription off this template produced an RNA denoted pool 130.1. Ligation of pool 130.1 to 30-P yielded Pool #1 (also referred to as LP160) shown in Figure 7C.

Seven cycles of PCR were performed according to published protocols (Ausubel et al., supra) with the following exceptions: (i) the starting concentration of RWR130.1 was 30 nanomolar, (ii) each primer was used at a concentration of 1.5 μM, (iii) the dNTP concentration was 400 μM for each base, and (iv) the Taq polymerase (Boehringer Mannheim) was used at 5 units per 100 μl. The double stranded product was purified on non-denaturing PAGE and isolated by electroelution. The amount of DNA was determined both by UV absorbance at 260 nm and ethidium bromide fluorescence comparison with known standards.

Enzymatic Synthesis of RNA. Transcription reactions from double stranded PCR DNA and synthetic oligonucleotides were performed as described previously (Milligan and Uhlenbeck, Meth. Enzymol. **180**:51 (1989)). Full length RNA was purified by denaturing PAGE, electroeluted, and desalted as described above. The pool RNA concentration was estimated using an extinction coefficient of 1300 O.D./μmole; RNA142, 1250 O.D./μmole; RNA 47.1, 480 O.D./μmole. Transcription from the double stranded pool DNA produced ~ 90 nanomoles of pool RNA.

Enzymatic Synthesis of RNA-Puromycin Conjugates. Ligation of the myc and pool messenger RNA sequences to the puromycin containing oligonucleotide was performed using a DNA splint, termed 19.35 (5'-TTT TTT TTT TAG CGC AAG A) (SEQ ID NO: 28) using a procedure analogous to that described by Moore and Sharp (Science **250**:992 (1992)). The reaction consisted of mRNA, splint, and puromycin olignucleotide (30-P, dA27dCdCP) in a mole ratio of 0.8 : 0.9 : 1.0 and 1-2.5 units of DNA ligase per picomole of pool mRNA. Reactions were conducted for one hour at room temperature. For the construction of the pool RNA fusions, the mRNA concentration was ~ 6.6 μmolar. Following ligation, the RNA-puromycin conjugate was prepared as described above for enzymatic reactions. The precipitate was resuspended, and full length fusions were purified on denaturing PAGE and isolated by electroelution as described above. The pool RNA concentration was estimated using an extinction coefficient of 1650 O.D./μmole and the myc template 1600 O.D./μmole. In this way, 2.5

nanomoles of conjugate were generated.

Preparation of dT_{25} Streptavidin Agarose. dT_{25} containing a 3' biotin was incubated at 1-10 μM with a slurry of streptavidin agarose (50% agarose by volume, Pierce, Rockford, IL) for 1 hour at room temperature in TE (10 mM Tris Chloride pH 8.2, 1 mM EDTA) and washed. The binding capacity of the agarose was then estimated optically by the disappearance of biotin-dT_{25} from solution and/or by titration of the resin with known amounts of complementary oligonucleotide.

Translation Reactions using E. coli Derived Extracts and Ribosomes. In general, translation reactions were performed with purchased kits (for example, E. coli S30 Extract for Linear Templates, Promega, Madison, WI). However, E. coli MRE600 (obtained from the ATCC, Rockville, MD) was also used to generate S30 extracts prepared according to published protocols (for example, Ellman et al., Meth. Enzymol. **202**:301(1991)), as well as a ribosomal fraction prepared as described by Kudlicki et al. (Anal. Biochem. **206**:389 (1992)). The standard reaction was performed in a 50 μl volume with 20-40 μCi of ^{35}S methionine as a marker. The reaction mixture consisted of 30% extract v/v, 9-18 mM $MgCl_2$, 40% premix minus methionine (Promega) v/v, and 5 μM of template (e.g., 43-P). For coincubation experiments, the oligos 13-P and 25-P were added at a concentration of 5 μM. For experiments using ribosomes, 3 μl of ribosome solution was added per reaction in place of the lysate. All reactions were incubated at 37°C for 30 minutes. Templates were purified as described above under enzymatic reactions.

Wheat Germ Translation Reactions. The translation reactions in Figure 8 were performed using purchased kits lacking methionine (Promega), according to the manufacturer's recommendations. Template concentrations were 4 μM for 43-P and 0.8 μM for LP77 and LP155. Reactions were performed at 25°C with 30 μCi ^{35}S methionine in a total volume of 25 μl.

Reticulocyte Translation Reactions. Translation reactions were performed either with purchased kits (Novagen, Madison, WI) or using extract prepared according to

published protocols (Jackson and Hunt, Meth. Enzymol. **96**:50 (1983)). Reticulocyte-rich blood was obtained from Pel-Freez Biologicals (Rogers, AZ). In both cases, the reaction conditions were those recommended for use with Red Nova Lysate (Novagen). Reactions consisted of 100 mM KCl, 0.5 mM MgOAc, 2 mM DTT, 20 mM HEPES pH 7.6, 8 mM creatine phosphate, 25 μM in each amino acid (with the exception of methionine if ^{35}S Met was used), and 40% v/v of lysate. Incubation was at 30°C for 1 hour. Template concentrations depended on the experiment but generally ranged from 50 nM to 1 μM with the exception of 43-P (Figure 6H) which was 4 μM.

For generation of the randomized pool, 10 ml of translation reaction was performed at a template concentration of ~ 0.1 μM (1.25 nanomoles of template). In addition, ^{32}P labeled template was included in the reaction to allow determination of the amount of material present at each step of the purification and selection procedure. After translation at 30°C for one hour, the reaction was cooled on ice for 30-60 minutes.

<u>Isolation of Fusion with dT_{25} Streptavidin Agarose.</u> After incubation, the translation reaction was diluted approximately 150 fold into isolation buffer (1.0 M NaCl, 0.1 M Tris chloride pH 8.2, 10 mM EDTA) containing streptavidin agarose (volume of slurry equal or greater than the volume of lysate) and incubated with agitation at 4°C for one hour. The agarose was then removed from the mixture either by filtration or centrifugation and washed with cold isolation buffer 2-4 times. The template was then liberated from the dT_{25} streptavidin agarose by repeated washing with 15 mM NaOH, 1 mM EDTA. The eluent was immediately neutralized in 3M NaOAc pH 5.2, 10 mM spermidine, and was ethanol precipitated. For the pool reaction, the total radioactivity recovered indicated approximately 50-70% of the input template was recovered.

<u>Isolation of Fusion with Thiopropyl Sepharose.</u> Fusions containing cysteine can be purified using thiopropyl sepharose as in Figure 13 (Pharmacia). In the experiments described herein, isolation was either carried out directly from the translation reaction or following initial isolation of the fusion (e.g., with streptavidin agarose). For samples purified directly, a ratio of 1:10 lysate to sepharose was used. For the pool, 0.5

ml of sepharose slurry was used to isolate all of the fusion material from 5 ml of reaction mixture. Samples were diluted into isolation buffer containing a slurry of thiopropyl sepharose and incubated with rotation for 1-2 hours at 4°C to allow complete reaction. The sepharose was washed repeatedly and recovered by centrifugation or filtration. The fusions were eluted from the sepharose using a solution of 25-30 mM dithiothreitol (DTT) in 10 mM Tris chloride pH 8.2, 1 mM EDTA. The fusion was then concentrated by a combination of evaporation under high vacuum and ethanol precipitation as described above. For the pool reaction, the total radioactivity recovered indicated approximately 1% of the template was converted to fusion.

Immunoprecipitation Reactions. Immunoprecipitations of peptide from translation reactions (Figure 10) were performed by mixing 4 μl of reticulocyte translation reaction, 2 μl normal mouse sera, and 20 μl Protein G + A agarose (Oncogene Science, Cambridge, MA; Calbiochem, San Diego, CA) with 200 μl of either PBS (58 mM Na_2HPO_4, 17 mM NaH_2PO_4, 68 mM NaCl), dilution buffer (10 mM Tris chloride pH 8.2, 140 mM NaCl, 0.025% NaN_3, 1% v/v Triton X-100), or PBSTDS (PBS + 1% Triton X-100, 0.5% deoxycholate 0.1% SDS). Samples were then rotated for one hour at 4°C, followed by centrifugation at 2500 rpm for 15 minutes. The eluent was removed, and 10 μl of c-myc monoclonal antibody 9E10 (Oncogene Science, Cambridge, MA) and 15 ul of Protein G + A agarose was added and rotated for 2 hours at 4°C. Samples were then washed with two 1 ml volumes of either PBS, dilution buffer, or PBSTDS. 40 μl of gel loading buffer (Oncogene Science Product Bulletin) was added to the mixture, and 20 μl was loaded on a denaturing PAGE as described by Schagger and von Jagow (Anal. Biochem. **166**:368 (1987)).

Immunoprecipitations of fusions (as shown in Figure 11) were performed by mixing 8 μl of reticulocyte translation reaction with 300 μl of dilution buffer (10 mM Tris chloride pH 8.2, 140 mM NaCl, 0.025% NaN_3, 1% v/v Triton X-100), 15 μl protein G sepharose (Sigma), and 10 μl c-myc antibody 9E10. After isolation, samples were treated with DNase free RNase A, labeled with polynucleotide kinase and ^{32}P gamma

ATP, and separated by denaturing PAGE (Figure 11).

Reverse Transcription of Fusion Pool. Reverse transcription reactions were performed according to the manufacturers recommendation for Superscript II, except that the template, water, and primer were incubated at 70°C for only two minutes (Gibco BRL, Grand Island, NY). 50 μCi alpha ^{32}P dCTP was included in some reactions to monitor extension.

Preparation of Protein G and Antibody Sepharose. Two aliquots of 50 μl Protein G sepharose slurry (50 % solid by volume) (Sigma) were washed with dilution buffer (10 mM Tris chloride pH 8.2, 140 mM NaCl, 0.025% NaN_3, 1% v/v Triton X-100) and isolated by centrifugation. The first aliquot was reserved for use as a precolumn prior to the selection matrix. After resuspension of the second aliquot in dilution buffer, 40 μg of c-myc AB-1 monoclonal antibody (Oncogene Science) was added, and the reaction incubated overnight at 4°C with rotation. The antibody sepharose was then purified by centrifugation for 15 minutes at 1500-2500 rpm in a microcentrifuge and washed 1-2 times with dilution buffer.

Selection. After isolation of the fusion and complementary strand synthesis, the entire reverse transcriptase reaction was used directly in the selection process. Two protocols are outlined here. For round one, the reverse transcriptase reaction was added directly to the antibody sepharose prepared as described above and incubated 2 hours. For subsequent rounds, the reaction is incubated ~2 hours with washed protein G sepharose prior to the antibody column to decrease the number of binders that interact with protein G rather than the immobilized antibody.

To elute the pool from the matrix, several approaches may be taken. The first is washing the selection matrix with 4% acetic acid. This procedure liberates the peptide from the matrix. Alternatively, a more stringent washing (e.g., using urea or another denaturant) may be used instead or in addition to the acetic acid approach.

PCR of Selected Fusions. Selected molecules are amplified by PCR using standard protocols as described above for construction of the pool.

USE OF IN VITRO PROTEIN SELECTION SYSTEMS

The selection systems of the present invention have commercial applications in any area where protein technology is used to solve therapeutic, diagnostic, or industrial problems. This selection technology is useful for improving or altering existing proteins as well as for isolating new proteins with desired functions. These proteins may be naturally-occurring sequences, may be altered forms of naturally-occurring sequences, or may be partly or fully synthetic sequences.

Isolation of Novel Binding Reagents. In one particular application, the RNA-protein fusion technology described herein is useful for the isolation of proteins with specific binding (for example, ligand binding) properties. Proteins exhibiting highly specific binding interactions may be used as non-antibody recognition reagents, allowing RNA-protein fusion technology to circumvent traditional monoclonal antibody technology. Antibody-type reagents isolated by this method may be used in any area where traditional antibodies are utilized, including diagnostic and therapeutic applications.

Improvement of Human Antibodies: The present invention may also be used to improve human or humanized antibodies for the treatment of any of a number of diseases. In this application, antibody libraries are developed and are screened in vitro, eliminating the need for techniques such as cell-fusion or phage display. In one important application, the invention is useful for improving single chain antibody libraries (Ward et al., Nature **341**:544 (1989); and Goulot et al., J. Mol. Biol. **213**:617 (1990)). For this application, the variable region may be constructed either from a human source (to minimize possible adverse immune reactions of the recipient) or may contain a totally randomized cassette (to maximize the complexity of the library). To screen for improved antibody molecules, a pool of candidate molecules are tested for binding to a target molecule (for example, an antigen immobilized as shown in Figure 2). Higher levels of stringency are then applied to the binding step as the selection progresses from

one round to the next. To increase stringency, conditions such as number of wash steps, concentration of excess competitor, buffer conditions, length of binding reaction time, and choice of immobilization matrix are altered.

Single chain antibodies may be used either directly for therapy or indirectly for the design of standard antibodies. Such antibodies have a number of potential applications, including the isolation of anti-autoimmune antibodies, immune suppression, and in the development of vaccines for viral diseases such as AIDS.

Isolation of New Catalysts. The present invention may also be used to select new catalytic proteins. In vitro selection and evolution has been used previously for the isolation of novel catalytic RNAs and DNAs, and, in the present invention, is used for the isolation of novel protein enzymes. This approach has two important advantages over catalytic antibody technology (reviewed in Schultz et al., J. Chem. Engng. News **68**:26 (1990)). First, in catalytic antibody technology, the initial pool is generally limited to the immunoglobulin fold; in contrast, the starting library of RNA-protein fusions may be either completely random or may consist of variants of known enzymatic structures. In addition, the isolation of catalytic antibodies generally relies on an initial selection for binding to transition state reaction analogs followed by laborious screening for active antibodies; again, in contrast, direct selection for catalysis is possible using an RNA-protein fusion library approach, as previously demonstrated using RNA libraries. In an alternative approach to isolating protein enzymes, the transition-state-analog and direct selection approaches may be combined.

Enzymes obtained by this method are highly valuable. For example, there currently exists a pressing need for novel and effective industrial catalysts that allow improved chemical processes to be developed. A major advantage of the invention is that selections may be carried out in arbitrary conditions and are not limited, for example, to in vivo conditions. The invention therefore facilitates the isolation of novel enzymes or improved variants of existing enzymes that can carry out highly specific transformations (and thereby minimize the formation of undesired byproducts) while functioning in

predetermined environments, for example, environments of elevated temperature, pressure, or solvent concentration.

An In Vitro Interaction Trap. The RNA-protein fusion technology is also useful for screening cDNA libraries and cloning new genes on the basis of protein-protein interactions. By this method, a cDNA library is generated from a desired source (for example, by the method of Ausubel et al., supra, chapter 5). To each of the candidate cDNAs, a peptide acceptor (for example, as a puromycin tail) is ligated (for example, using the techniques described above for the generation of LP77, LP155, and LP160). RNA-protein fusions are then generated as described herein, and the ability of these fusions (or improved versions of the fusions) to interact with particular molecules is then tested as described above. If desired, stop codons and 3' UTR regions may be avoided in this process by either (i) adding suppressor tRNA to allow readthrough of the stop regions, (ii) removing the release factor from the translation reaction by immunoprecipitation, (iii) a combination of (i) and (ii), or (iv) removal of the stop codons and 3' UTR from the DNA sequences.

The fact that the interaction step takes place in vitro allows careful control of the reaction stringency, using nonspecific competitor, temperature, and ionic conditions. Alteration of normal small molecules with non-hydrolyzable analogs (e.g., ATP vs. ATPgS) provides for selections that discriminate between different conformers of the same molecule. This approach is useful for both the cloning and functional identification of many proteins since the RNA sequence of the selected binding partner is covalently attached and may therefore be readily isolated. In addition, the technique is useful for identifying functions and interactions of the ~50-100,000 human genes, whose sequences are currently being determined by the Human Genome project.